Manipulating Data (Linear Transformations)

Use the list "PRES" that you transferred the other day to complete the following.

1. Create a histogram of the data on your calculator. Briefly describe the distribution (you do not need to draw a picture here). Put the data into L1.

2. Find the mean, median, IQR and standard deviation of this original set of data. (1-VAR STATS)

x = ____________ M = ____________ s = ____________ IQR = ____________

3. Now take the data and **add 30** to each observation. *(Hint: use your lists to do the calculations for you! Put the data into L2)* Find the mean, median, IQR, and standard deviation of this new set of data. Also look at the histogram of this new set of data on your calculator.

$\bar{x}$ = 99.5 M = 97 s = 11.88 IQR = 18

4. Which statistics changed? How? How does the histogram compare to the original histogram?

center
+30

Shape: same
Center: ↓

5. Now take the **original** set of data and **multiply each observation by 2.5**. *(put the data into L3)* Find the mean, median, IQR, and standard deviation of this new set of data. Also look at the histogram of this new set of data on your calculator.

$\bar{x}$ = 173.7 M = ________ s = ________ IQR = ________

6. Which statistics changed? How? How does the histogram compare to the original histogram?

center & spread
x 2.5
shape: same

7. Now take the **original** set of data and multiply each observation by 5 and add 45. Find the mean, median, IQR, and standard deviation of this new set of data. Also look at the histogram of this new set of data on your calculator.

$\bar{x}$ = 392.4 M = ________ s = ________ IQR = ________

×5 +45 ×5

8. Which statistics changed? How? How does the histogram compare to the original histogram?

×5 +45

shape same

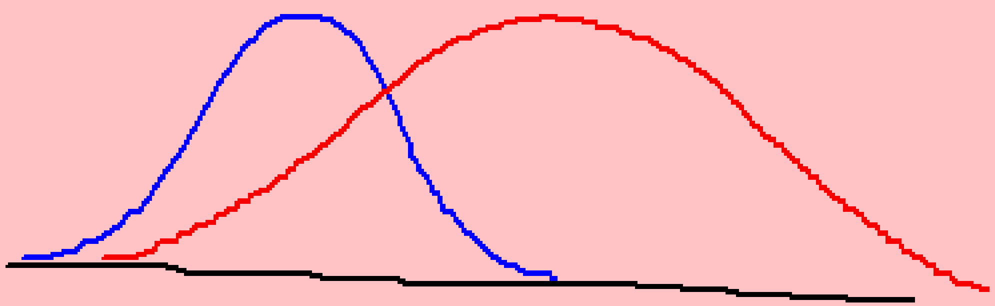
Manipulating Data (Linear Transformations)

- <u>Multiplying</u> an entire data set by a constant "*b*" changes….

  center + spread

  - How? $\times b$

  - Examples:



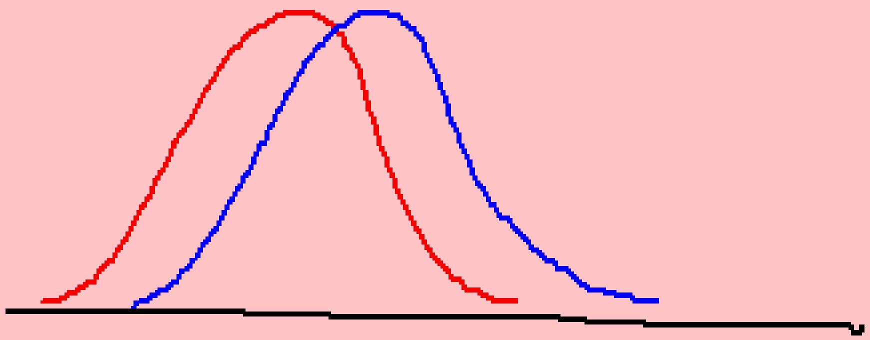
	Data	$\times 3$
$\bar{x}$	5	15
M	7	21
s	2	6
IQR	4	12

$\times \frac{1}{2}$

- Adding a constant "*a*" to an entire data set changes…

  center

  - How? $+a$
  - Examples: $+2''$



$\bar{x} = 65''$ $\bar{x} = 67''$

$s = 2''$ $s = 2''$

Example:

I have a distribution with the following statistics:

$\bar{x} = 25.3$
M = 21
s = 3.1
IQR = 8

$\times 4 \quad +6$

If I multiply each observation by 4 and add 6, what will the new Statistics be?

$\bar{x} =$ 107.2

M = 90

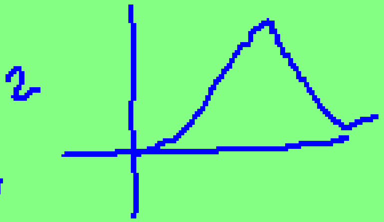
s = 12.4

IQR = 32

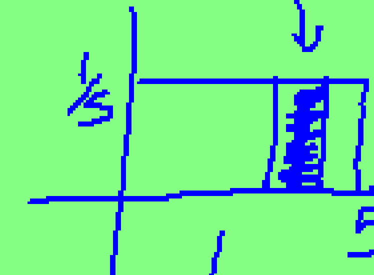
AP STAT: Section 1.3

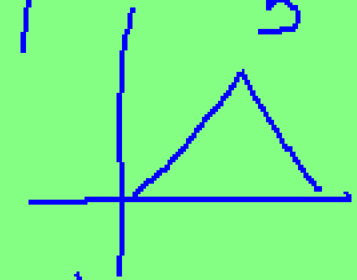
Other Distributions (using mean and std. deviation)

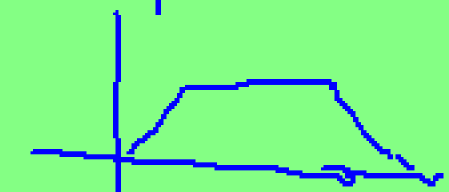
graph/picture



0.2







Histogram/stemplot/etc.	**Density Curve**
• use samples	• population
• actual obs.	• smooth curve
• diff. types	• idealized model
	• like to rel. freq. hist. show % - area = 1
	• area = %

Use:

$\overline{X}$ = sample mean

S = sample std. dev.

0.4

use

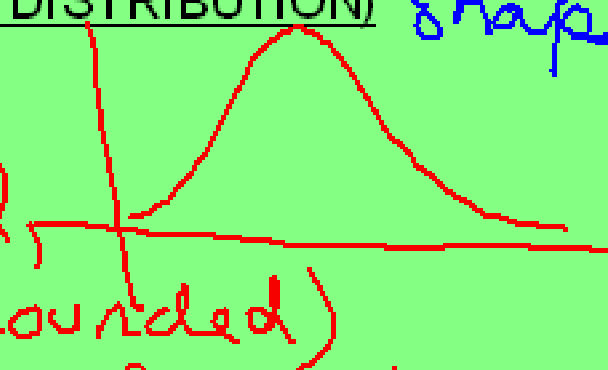
μ = pop. mean

σ = pop. std. dev.

* many shapes

Specific Density Curve:

<u>NORMAL CURVE (or DISTRIBUTION)</u>



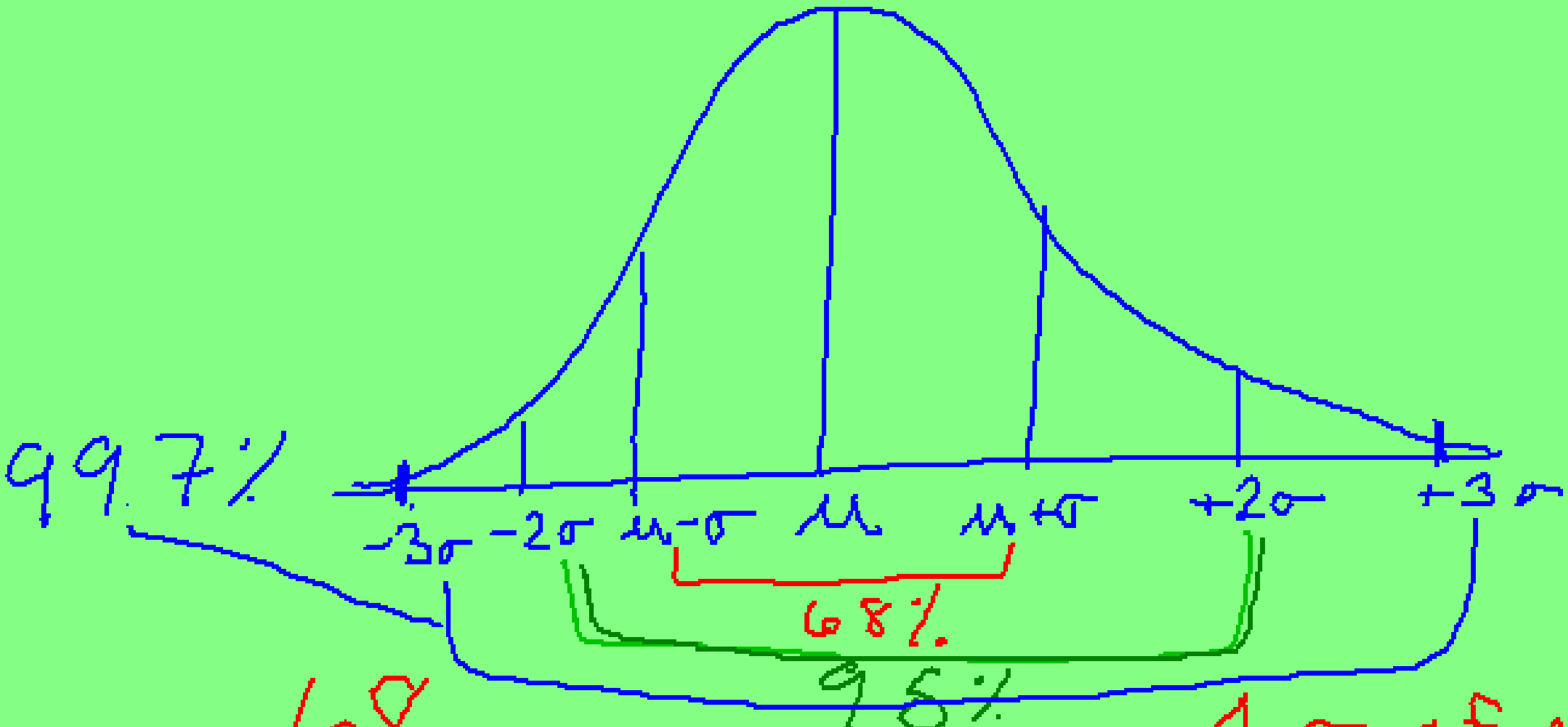
• sym, unimodal, bell-shaped (mounded)

• describe "normal" data

• $N(\mu, \sigma)$ $N(65, 5)$

Empirical Rule

In a normal distribution with N(μ, σ)….



- <u>68</u> % of the observations fall within <u>1σ of μ</u> ($\mu \pm \sigma$)
- <u>95</u> % of the observations fall within <u>2σ of μ</u> ($\mu \pm 2\sigma$)
- <u>99.7</u> % of the observations fall within <u>3σ of μ</u> ($\mu \pm 3\sigma$)

Example:

The avg. time that it takes a crew from Sweeper Sam to clean a room is normally distributed with a mean of 2.1 hrs and a standard deviation of 0.3 hrs.

$N(2.1, 0.3)$

a) The total clean up time will fall within what interval 95% of the time?

$2.1 + 2(0.3) = 1.5$

$2.1 - 2(0.3) = 2.7$

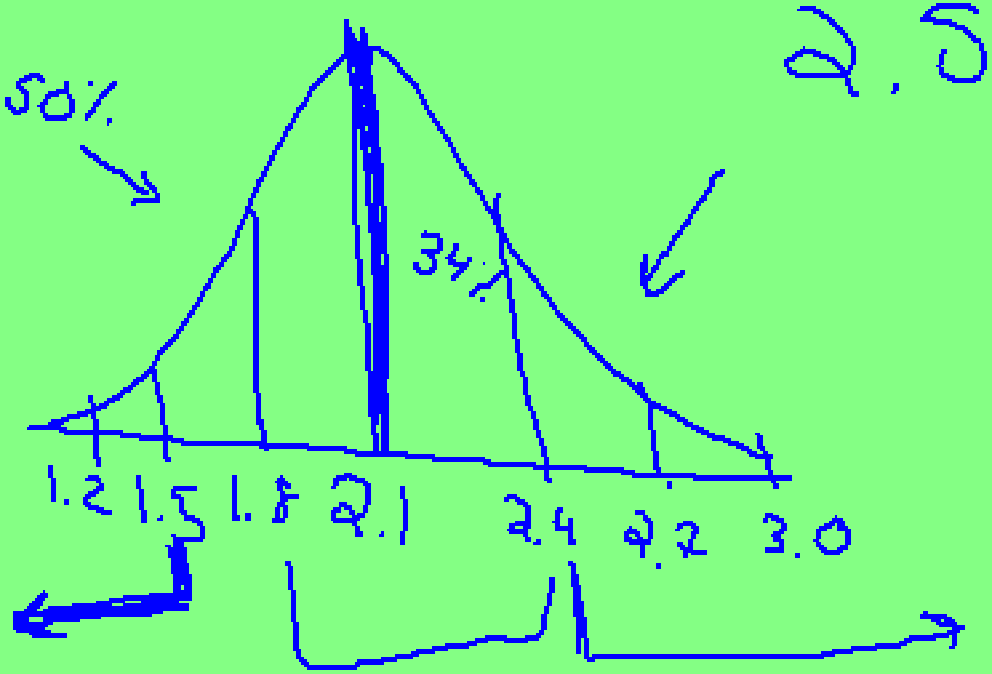
$\mu \pm 2\sigma$

b) What proportion of the time will it take the crew ~~2.7~~ 2.4 hours or more?

16%

c) What percent of the time will it take the crew 1.5 hrs or less?

2.5%



Standardizing Observations

1) You are in a history class and a math class. You take your chapter 1 test in both classes on the same day, and get them both back a few days later. Your grades are as follows. Which class did you do better in?

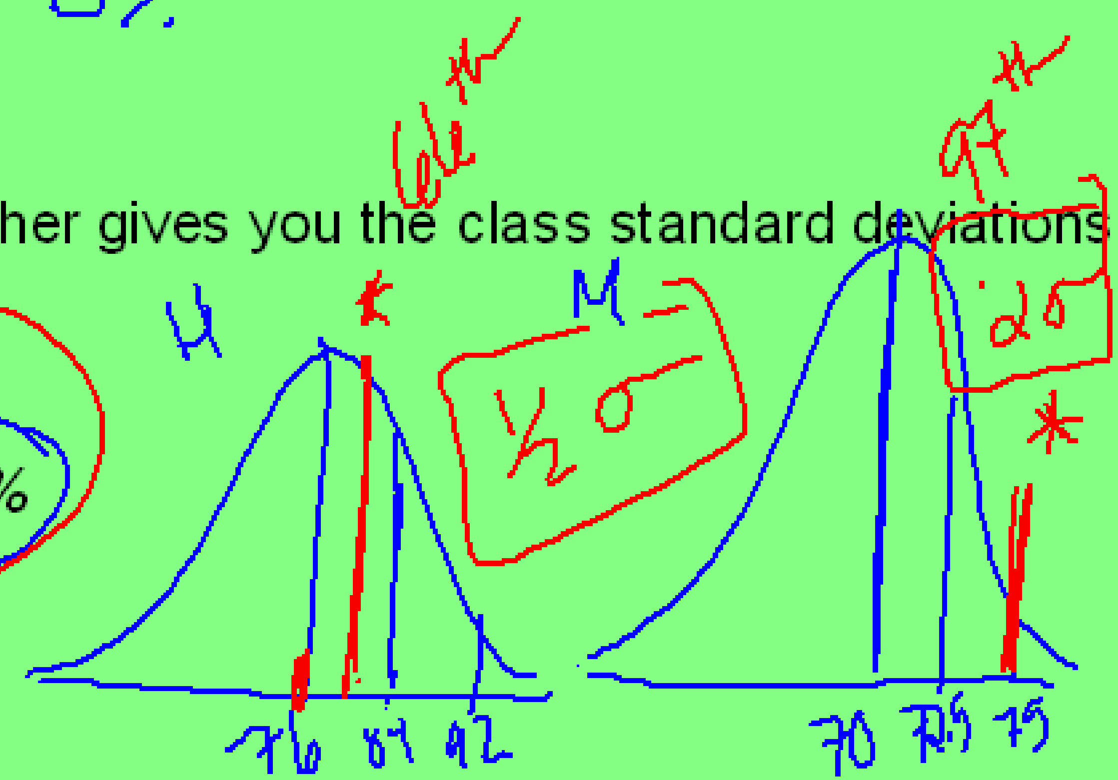
History: 81% *Math: 75%*

2) Same scenario. However, each teacher tells you a bit about how your class did. Which class did you do better in?

+5% < *History:* 81 *Math:* 75 > 5%
mean: 76% *mean:70%*

3) Same scenario. However how your teacher gives you the class standard deviations. Which class did you do better in?

History: *Math:*
std. dev: 8% *std. dev:2.5%*



Standardizing Observations – Z-scores

Question: *How can we compare one distribution to another distribution if they don't have the same parameters (mean and std. dev)?*

W = 68" μ = 65"

M = 68" μ = 67"

Answer:

comparing the obs. to their μ + σ

To standardize: (Z-scores)

- measure observations…. in terms of how far they are above (or below) their μ and σ.
- Z = $\dfrac{X - \mu}{\sigma}$

obs → X; its μ & σ

Z = 1.2 σ's above μ

- Z- score tells us… how far an obs. is away from μ in terms of σ

\+ Z-score = above μ

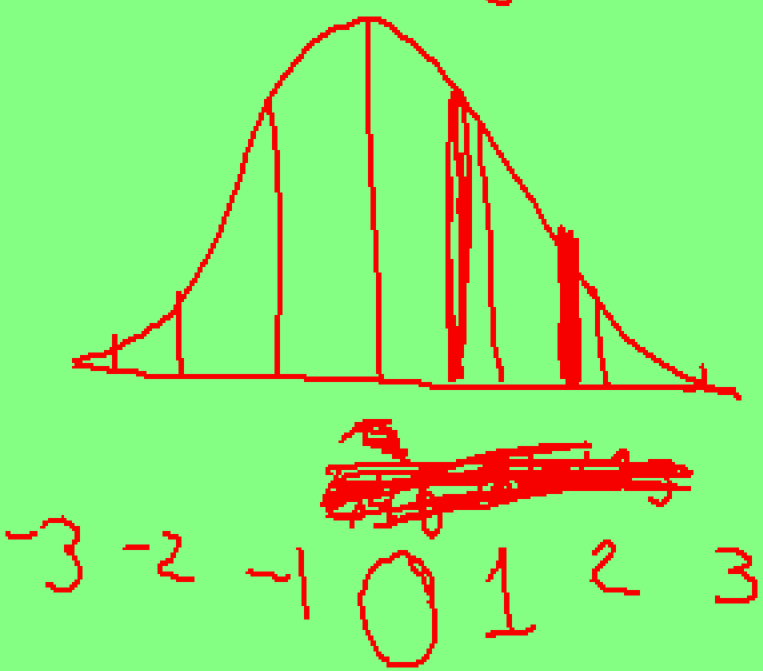
\- Z-score = below μ

Example: *The heights of 18-24 year old women are normally distributed with the following:*

mean = 64.5" and std. dev = 2.5" $N(64.5, 2.5)$

We know a woman who is 69" tall. How does she compare to the rest of the women in her age group? Calculate her z score, and interpret what it means.

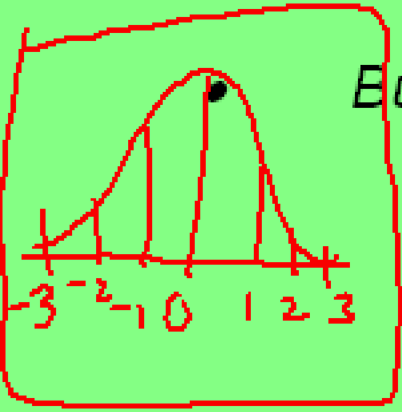
$$z = \frac{x-\mu}{\sigma} = \frac{69-64.5}{2.5} = 1.80\,\sigma\text{'s above her mean}$$



inches to ft.

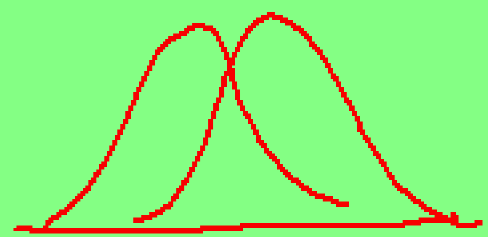
Notes on Standardizing a distribution:

- Standardizing one observation… compares it to its mean
- But we can also standardize… a whole distribution, by comparing the whole dist. to its μ + σ



- But what does this do to the shape of the distribution?

  Same



- Standardizing is actually just…

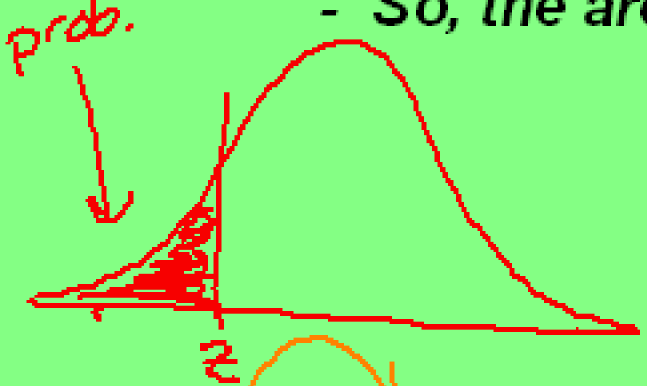
  $-\mu$ and $\div\sigma$

  *just like Ex: PRES

  +30, ×2.5

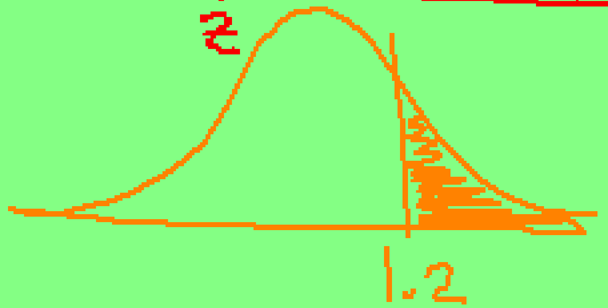
- When we do this we have a new distribution, called: Standard Norm. Distr.

- Table A in the book
  - Gives: Std. normal distr., gives the prob. (%)
  - So, the area to the left of the z-score represents… BELOW z-scores



$P(z<1.8) = 0.9641$

96.41%

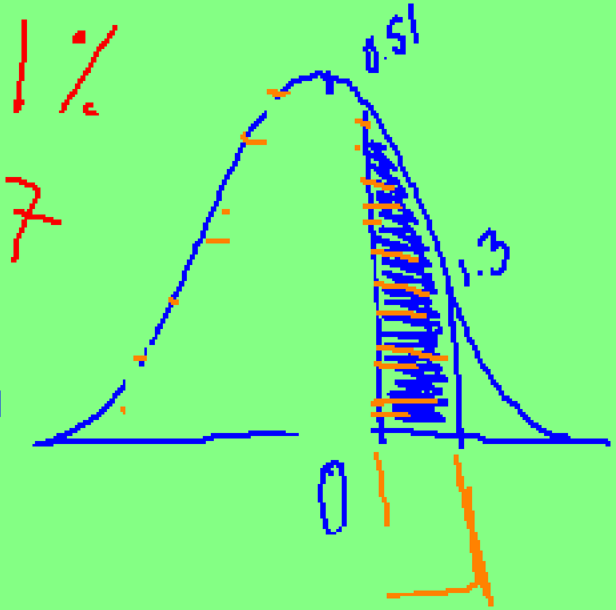


$P(z<2.83) = 0.9977$

$P(z>1.2) = 11.51\%$

$P(0.51<z<1.3)$

0.2082



Back to the height example….

Remember that the heights of 18-24 year old women are $\boxed{N(64.5”, 2.5”)}$. What percentile is the girl who is 68” tall?

shape
center
spread

$P(x < 68") \rightarrow z = \dfrac{68 - 64.5}{2.5} = 1.4 \quad P(z < 1.4) = 0.9192$

91.92^th^

What percent of 18-24 year old women are less than 5 feet tall?

$P(x < 60") \rightarrow z = \dfrac{60 - 64.5}{2.5} = -1.8 \quad P(z < -1.8) = 3.59\%$

What percent 18-24 year old of women are over 5'8” tall?

$P(x > 68") = 1 - 0.9192 = 8.08\%$

** ***PROBABILITY NOTATION!!*** $P(X \geq \#)$ $P(z \geq \_\_\_)$

Another example:

Blood pressures of high school students are N(170, 30). What is the probability that a randomly selected high school student has a blood pressure of 180 or above?

$P(X \geq 180) \rightarrow \cancel{z = 0.33 \quad P(z \geq 0.33)} = 0.3707$

Using the same data as above, what is the probability that a high school student has a blood pressure between 160 and 230?

$P(160 < X < 230)$

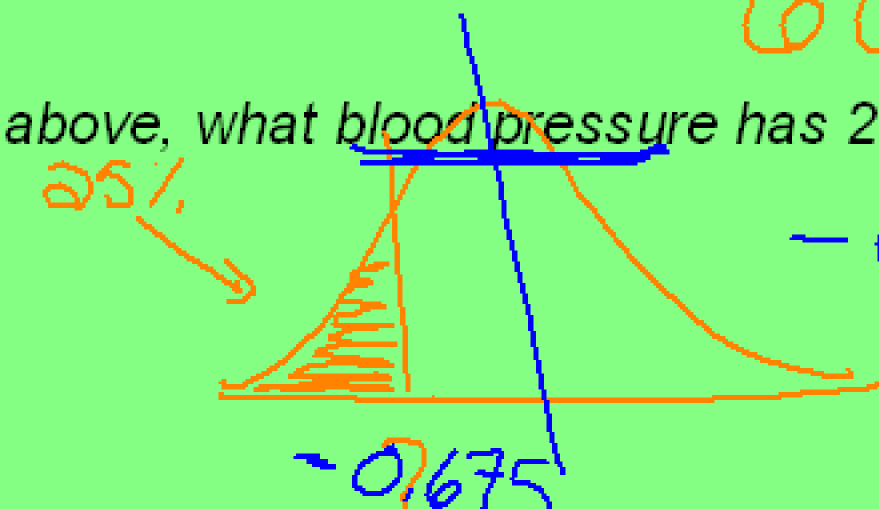
230 $z = 2$

160 $z = -0.33$

60.65%

Using the same data as above, what blood pressure has 25% of the bobservations below it?

$P(z > 1.2)$



$-0.675 = \frac{X - 170}{30}$

Calculator use:

To find the percent of observations between 2 points:

normalcdf(lower bound, upper bound, μ, σ) = %

(a: lower bound, b: upper bound)

$P(a < X < b)$

To find what observation has a certain percent of the data below it:

invnorm(prop. below, μ, σ)

(0.25 above "prop. below")

$P(X > 180) = P(180 < X < \text{inf.})$

* ***On the calculator, infinity is:***

$E99 = \infty$

$-\infty = -E99$